

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Wallis, Nicola G.
Burnham, Martin K. R.

(ii) TITLE OF INVENTION: murC

(iii) NUMBER OF SEQUENCES: 6

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Dechert, Price & Rhoads
(B) STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
(C) CITY: Philadelphia
(D) STATE: PA
(E) COUNTRY: USA
(F) ZIP: 19103-2793

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: Windows 95
(D) SOFTWARE: FastSEQ for Windows Version 2.0b

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: 60/052,720
(B) FILING DATE: 03-JUL-1997

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Falk, Stephen T

09103287 062398

(B) REGISTRATION NUMBER: 36,795
(C) REFERENCE/DOCKET NUMBER: GM10025

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 215-994-2488
(B) TELEFAX: 215-994-2222
(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1351 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATGAGTAAGG AGTTTTATAT AATGACACAC TATCATTTTG TCGGAATTAA AGGTTCTGGC 60
ATGAGTTCAT TAGCACAAAT CATGCATGAT TTAGGACATG AAGTTCAAGG ATCGGATATT 120
GAGAACTACG TATTTACAGA AGTTGCTCTT AGAAATAAGG GGATAAAAAT ATTACCATTT 180
GGTGCTAATA ACATAAAGA AGATATGGTA GTTATACAAG GTAATGCATT CGCGAGTAGC 240
CATGAAGAAA TAGTACGTGC ACATCAATTG AAATTAGATG TTGTAAGTTA TAATGATTTT 300
TTAGGACAGA TTATTGATCA ATATACTTCA GTAGCTGTAA CTGGTGCACA TGGTAAAACT 360
TCTACAACAG GTTTATTATC ACATGTTATG AATGGTGATA AAAAGACTTC ATTTTAAATT 420
GGTGATGGCA CAGSTATGGG ATTGCTTGAA AGTGATTATT TCGCTTTTGA GGCATGTGAA 480
TATAGACGTC ACITTTTAAAG TTATAAACCT GATTACGCAA TTATGACAAA TATTGATTTC 540
GATCATCCTG ATTATTTCAA AGATATTAAT GATGTTTTTG ATGCATTCCA AGAAATGGCA 600
CATAATGTTA AAAAAGGTAT TATTGCTTGG GGTGATGATG AACATCTACG TAAAATTGAA 660
GCAGATGTTT CAATTTATTA CTATGGATTT AAAGATTCGG ATGACATTTA TGCTCAAAAT 720
ATTCAAATTA CGGATAAAGG TACTGCTTTT GATGTGTATG TGGATGGTGA GTTTTATGAT 780
CACTTCCTGT CTCCACAATA TGGTGACCAT ACAGTTTAA ATGCATTAGC TGTAATTGCG 840
ATTAGTTATT TAGAGAAGCT AGATGTTACA AATATTAAAG AAGCATTAGA AACGTTTGGT 900
GGTGTTAAAC GTCGTTTCAA TGAAACTACA ATTGCAAATC AAGTTATTGT AGATGATTAT 960
GCACACCATC CAAGAGAAAT TAGTGCTACA ATTGACACAG CACGAAAGAA ATATCCACAT 1020
AAAGAAGTTG TTGCAGTATT TCAACCACAC ACTTCTCTA GAACACAAGC ATTTTAAAT 1080
GAATTTCAG AAAGTTTATG TAAAGCAGAT CGTGATTCT TATGTGAAAT TTTTGGCTCA 1140
ATTAGAGAAA ATTCTGGCGC ATTAACGATA CAAGATTTAA TTGATAAAAT TGGAGGTGCA 1200
TCGTTTATTA ATGAAGATCT TATTAATGTA TTAGAACAAT TTGATAATGC TGTTGTTTTA 1260

TTTATGGGTG CAGGTGATAT TCAAAAATTA CAAAATGCAT ATTTAGATAA ATTAGGCATG 1320
 AAAAATGCGT TTTAATATGT TTATAATAGA G 1351

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 437 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Thr His Tyr His Phe Val Gly Ile Lys Gly Ser Gly Met Ser Ser
 1 5 10 15
 Leu Ala Gln Ile Met His Asp Leu Gly His Glu Val Gln Gly Ser Asp
 20 25 30
 Ile Glu Asn Tyr Val Phe Thr Glu Val Ala Leu Arg Asn Lys Gly Ile
 35 40 45
 Lys Ile Leu Pro Phe Gly Ala Asn Asn Ile Lys Glu Asp Met Val Val
 50 55 60
 Ile Gln Gly Asn Ala Phe Ala Ser Ser His Glu Glu Ile Val Arg Ala
 65 70 75 80
 His Gln Leu Lys Leu Asp Val Val Ser Tyr Asn Asp Phe Leu Gly Gln
 85 90 95
 Ile Ile Asp Gln Tyr Thr Ser Val Ala Val Thr Gly Ala His Gly Lys
 100 105 110
 Thr Ser Thr Thr Gly Leu Leu Ser His Val Met Asn Gly Asp Lys Lys
 115 120 125
 Thr Ser Phe Leu Ile Gly Asp Gly Thr Gly Met Gly Leu Pro Glu Ser
 130 135 140
 Asp Tyr Phe Ala Phe Glu Ala Cys Glu Tyr Arg Arg His Phe Leu Ser
 145 150 155 160
 Tyr Lys Pro Asp Tyr Ala Ile Met Thr Asn Ile Asp Phe Asp His Pro
 165 170 175
 Asp Tyr Phe Lys Asp Ile Asn Asp Val Phe Asp Ala Phe Gln Glu Met
 180 185 190
 Ala His Asn Val Lys Lys Gly Ile Ile Ala Trp Gly Asp Asp Glu His
 195 200 205
 Leu Arg Lys Ile Glu Ala Asp Val Pro Ile Tyr Tyr Tyr Gly Phe Lys

210 215 220
 Asp Ser Asp Asp Ile Tyr Ala Gln Asn Ile Gln Ile Thr Asp Lys Gly
 225 230 235 240
 Thr Ala Phe Asp Val Tyr Val Asp Gly Glu Phe Tyr Asp His Phe Leu
 245 250 255
 Ser Pro Gln Tyr Gly Asp His Thr Val Leu Asn Ala Leu Ala Val Ile
 260 265 270
 Ala Ile Ser Tyr Leu Glu Lys Leu Asp Val Thr Asn Ile Lys Glu Ala
 275 280 285
 Leu Glu Thr Phe Gly Gly Val Lys Arg Arg Phe Asn Glu Thr Thr Ile
 290 295 300
 Ala Asn Gln Val Ile Val Asp Asp Tyr Ala His His Pro Arg Glu Ile
 305 310 315 320
 Ser Ala Thr Ile Asp Thr Ala Arg Lys Lys Tyr Pro His Lys Glu Val
 325 330 335
 Val Ala Val Phe Gln Pro His Thr Phe Ser Arg Thr Gln Ala Phe Leu
 340 345 350
 Asn Glu Phe Ala Glu Ser Leu Cys Lys Ala Asp Arg Val Phe Leu Cys
 355 360 365
 Glu Ile Phe Gly Ser Ile Arg Glu Asn Ser Gly Ala Leu Thr Ile Gln
 370 375 380
 Asp Leu Ile Asp Lys Ile Gly Gly Ala Ser Phe Ile Asn Glu Asp Leu
 385 390 395 400
 Ile Asn Val Leu Glu Gln Phe Asp Asn Ala Val Val Leu Phe Met Gly
 405 410 415
 Ala Gly Asp Ile Gln Lys Leu Gln Asn Ala Tyr Leu Asp Lys Leu Gly
 420 425 430
 Met Lys Asn Ala Phe
 435

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 660 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATTTAAAGAT TCGGATGACA TTTATGCTCA AATATTTCAA ATTACGGATA AAGGTACTGC 60
 TGTTGATGTG TATGTGGATG GTGAGTTTTA TGATCACTTC CTGTCTCCAC AATATGGTGA 120
 CCATACAGTT TTAAATGCAT TAGCTGTAAT TGCGATTAGT TATTTAGAGA AGCTAGATGT 180
 TACAAATATT AAAGAAGCAT TAGAAACGTT TGGTGGTGTT AAACGTCGTT TCAATGAAAC 240
 TACAATTGCA AATCAAGTTA TTGTAGATGA TTATGCACAC CATCCAAGAG AAATTAGTGC 300
 TACAATTGAC ACAGCACGAA AGAAATATCC ACATAAAGAA GTTGTTGCAG TATTTCAACC 360
 ACACACTTTC TCTAGAACAC AAGCATTTTT AAATGAATTT GCAGAAAGTT TAGTAAAGC 420
 AGATCGTGTA TTCTTATGTG AAATTTTTGG ATCAATTAGA GAAAATACTG GCGCATTAAAC 480
 GATACAAGAT TTAATTGATA AAATTGAAGG TGCATCGTTA ATTAATGAAG ATTCTATTAA 540
 TGTATTAGAA CAATTTGATA ATGCTGTTGT TTTATTTATG GGTGCAGGTG ATATTCAAAA 600
 ATTACAAAT GCATATTTAG ATAAATTAGG CATGAAAAAT GCGTTTAAAT ATGTTTATAA 660

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 215 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Phe	Lys	Asp	Ser	Asp	Asp	Ile	Tyr	Ala	Gln	Ile	Phe	Gln	Ile	Thr	Asp
1				5					10					15	
Lys	Gly	Thr	Ala	Val	Asp	Val	Tyr	Val	Asp	Gly	Glu	Phe	Tyr	Asp	His
			20					25					30		
Phe	Leu	Ser	Pro	Gln	Tyr	Gly	Asp	His	Thr	Val	Leu	Asn	Ala	Leu	Ala
			35				40					45			
Val	Ile	Ala	Ile	Ser	Tyr	Leu	Glu	Lys	Leu	Asp	Val	Thr	Asn	Ile	Lys
	50					55				60					
Glu	Ala	Leu	Glu	Thr	Phe	Gly	Gly	Val	Lys	Arg	Arg	Phe	Asn	Glu	Thr
65					70				75					80	
Thr	Ile	Ala	Asn	Gln	Val	Ile	Val	Asp	Asp	Tyr	Ala	His	His	Pro	Arg
			85					90					95		
Glu	Ile	Ser	Ala	Thr	Ile	Asp	Thr	Ala	Arg	Lys	Lys	Tyr	Pro	His	Lys
		100						105					110		
Glu	Val	Val	Ala	Val	Phe	Gln	Pro	His	Thr	Phe	Ser	Arg	Thr	Gln	Ala
		115					120					125			
Phe	Leu	Asn	Glu	Phe	Ala	Glu	Ser	Leu	Ser	Lys	Ala	Asp	Arg	Val	Phe
	130						135					140			

SECRET

(i) SEQUENCE CHARACTERISTICS:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

19

(i) ~~SEQUENCE~~ CHARACTERISTICS:

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

19